ROBO-GUIDE: towards safe, reliable, trustworthy, and natural behaviours in robotic assistants

Jonathan M. Aitken, Luke Boorman, David Cameron, Adriel Chua, Emily C. Collins, Samuel Fernando, James Law, Uriel Martinez-Hernandez, Owen McAree

Sheffield Robotics, The University of Sheffield, Sheffield, UK
{jonathan.aitken,o.mcaree,l.boorman,d.s.cameron,dxachua1, e.c.collins,s.fernando,j.law,uriel.martinez}@sheffield.ac.uk

Abstract. In this paper we introduce our approach to developing a robotic guide, taking a broad view to account for advancements in safety, verification, behavioural psychology, natural language, and cognitive neuroscience. We describe a novel scenario, whereby robotic assistants are required to ask for assistance to operate a lift, and results from a preliminary investigation into floor verification using readily-available information. The aim being to create assistive robots that can naturally integrate into existing infrastructure.

1 Introduction

For assistive robots (in roles such as carers, guides, companions, and assistants) to be most effective, they will need to seamlessly integrate into the human-centric environments we have designed, interacting through natural communication methods, whilst being safe, reliable, and trustworthy. The ROBO-GUIDE (ROBOtic GUidance and Interaction DEvelopment) project is an interdisciplinary project bringing together engineers and scientists working in computational neuroscience, control systems, formal verification, natural language, and psychology, to address how such a system can be designed and built with a view of complete system integration. The aim of this project is to develop a guide robot that can navigate inside a large working building, filled with people who are not, on the whole, familiar with robotic technology, and to do so in a safe, and reliable way. As the robot is relied upon to assist others, it is also important that interactions are natural and engender trust.

A novel scenario, which we are investigating in our current work, is how the robot navigates between floors using a lift. This is a particularly interesting, multifaceted challenge, in which the importance of an integrated approach is exacerbated. The robot is required to: navigate into and out of the lift safely, and in a timely manner, without causing injury or damage; interact through natural means with other lift users to request assistance in controlling the lift in a way that elicits assistance; operate in a trustworthy and reassuring manner around others in the closed and close-proximity environment; reliably identify floors in the building in order to correctly navigate to its destination, using existing information designed for human navigation; recall other users and tailor its interactions to further engender trust.

These issues all have relevance, in some form, to the wider field of assistive robotics, including for guides, couriers, carer, and companion robots. For example, Rosenthal et al. [13] show the importance of a symbiotic human-robot relationship, whereby a robot that asks for assistance from a human can complete its role more efficiently; Dixon et al. [5], highlight the need for formal verification in assistive robots to ensure safe operation, and engender trust; and Kulyukin [8] examines the appropriateness of natural language as an interface for interaction with assistive robots, and identifies the potential for its use in partially autonomous robots requiring some human intervention.

In the remainder of this paper, we describe these key concepts in more detail, and present initial results from an experiment to investigate how confidently the robot can identify a particular floor within a building, from a combination of readily-available human and robotic navigational information.

2 Considerations for a guide robot

For this application, we have selected to use the Pioneer LX from Adept MobileRobots[1]. This is an extendible, wheeled platform with a 13 hour run time, laser scanner for indoor mapping and navigation, proximity and impact sensors, speech synthesis software, and autonomous charging capacity, which lends itself well to the application. In the following sections we outline some of the major considerations for the robot-guide task, with particular reference to this platform, and our approaches to them.

2.1 Navigation

In order for a robot to navigate within a building, it has to be able to map the building, locate itself on the map and avoid obstacles while moving. However, these are well-researched issues, and so not the focus of our attention. We have used the inbuilt mapping, navigation, and obstacle avoidance software provided with the Pioneer LX.

2.2 Safety

Any system that is deployed in and around members of the general public must be verifiably safe in its operation. In order to show that a system is safe to operate within a given environment a safety case can be constructed [10], which produces an argument that demonstrates why a system is safe to operate in those particular conditions; this can be presented using Goal Structured Notation (GSN) [7].

[1] http://www.mobilerobots.com/ResearchRobots/PioneerLX.aspx

In the case of a robot guide, it is especially important to be able to argue why the robot is safe: it will lead guests through the building and have to navigate past other, potentially uncooperative, members of the public. In order to understand the robot, we must understand the boundaries of operation, to understand how it must behave within the environment. This will involve developing an understanding about the operational requirements of the robot.

In the intended application, a key concern for safety will surround the robot's interaction with humans, especially the general public. As such, we are developing safety patterns for use in the design of such systems [1].

2.3 System verification

To perform its task, the robot will need to make a number of decisions about how to act in the environment. It is important that these decisions be verified against a specification of how the robots is expected to perform [1]. This specification needs to be derived from the perspective of both operational safety and task performance.

The previous section introduced the safety challenges faced by the robot, particularly with regard to its close proximity to unsuspecting humans. Once a software suite is deployed to the robot a significant amount of testing is required to ensure it performs safely. However, due to the complex and unpredictable nature of the environment it is not possible to perform a fully exhaustive set of tests. Therefore, in addition to physical tests, it is necessary to formally verify the decision making of the robot to ensure that it will never intentionally act in an unsafe manner.

Once safe operation of the robot can be guaranteed, attention must turn to its performance of the task. A particular challenge faced by the robot is the need to use a lift to change floors. This is made even more complex by the fact that the robot cannot operate the lift itself, but must ask for human assistance. It is perfectly plausible, therefore, that the robot finds itself on an incorrect floor. Once again, it is necessary to formally verify the decision making of the robot to ensure that if this happens, it is able to recover the situation.

It is important to consider both safety and performance when verifying the decision making of the robot as otherwise it is possible for one to dominate. For example, a perfectly safe robot may simply decide to park itself in the corner of a room so as not to pose a hazard to anyone, but of course it will likely never achieve its goal[2]. Conversely, a goal driven robot may drive the fastest route to the goal without stopping for anyone, but this is likely to cause harm to at least one person! The perfect robot would achieve a balance somewhere between these two.

2.4 Human-robot interaction: promoting helping behaviour

In addition to safely operating around humans, a guide robot is also required to appropriately interact with them. Whilst the core application is to assist other

[2] with the exception of the goal *park in the corner of the room*

building users, it will also need to request assistance, for example to open doors or operate the lift. How will the robot get help from humans in these situations? How will the ambiguity involved in these scenarios be overcome in order that humans near the robot understand that it requires help, and more importantly feel comfortable and safe helping it? To begin to answer these questions we explore insights from social psychology, focusing on factors of user trust and situational ambiguity.

Successful interpersonal cooperation identifies trust as its foundation [9], as the successful completion of tasks involving two agents requires that both individuals trust each other. This human-human theory of cooperation has analogues in trust promotion within HRI [6]. We are aware that interaction with robots in cooperative environments may present a novel scenario for many people, and from this potential ambiguity could arise in which the human is unsure how to approach or assist the robot. Ambiguity in helping scenarios results in substantial detriment to pro-active helping behaviour [4]. In order that our robot be aided most effectively by humans around it we must consider just how trustworthy our robot is perceived to be in order that not only is helping behaviour freely given by human agents, but also that such helping behaviour is given comfortably by the human. The ROBO-GUIDE platform will therefore be required to communicate a clear plan or intention of its task to reduce ambiguity, as well as allow a human to understand in as clear a way as possible its limitations and its requirements for help. As such the robot will be designed to influence its perceived trustworthiness using a combination of affective and cognitive trust markers [9] in the language it uses to convey its intentions; for example by using friendly statements, and clearly communicating its aims, *Please follow me; I am here as your tour guide* [3].

2.5 Natural language for interacting with mobile robots

To facilitate such human-robot interaction, we chose to use speech as the most natural form of human communication. Although not always the ideal interface for interacting with robots there are a number of advantages for speech as a communication channel: it requires no special training, is high bandwidth, and hands-free, all of which make it attractive for control applications [2]. Previous work on natural language and mobile robots has mainly focussed on command systems where users can convey instructions for the robot to perform. However, in our project the robot has a different role: instead of receiving commands and instructions, the robot should use language spontaneously and naturally as part of navigating around a busy working environment. Ideally the robot should interact with humans in the vicinity to assist with tasks such as opening doors or operating the lift.

In addition to using speech to communicate with other building users, there are situations wherein the robot can use speech to communicate with other devices via existing human-machine interfaces. One such example is the lift used in our current experiments, which announces both the floor and direction of travel. In our initial experiments we train a speech recognition system to recognise the announcements given by the lift. We describe the speech recognition system setup that we use and present evaluation results in Section 3.2.

2.6 Memory

During the course of its role, the robot is likely to encounter the same individuals on a regular basis. To enable it to tailor its actions, the robot should have the ability to learn about these individuals and recall pertinent information. This will help raise efficiency, and possibly improve the interactions discussed in sections 2.4 and 2.5. In related work we are investigating human memory systems, which allow the efficient storage and recall of past events and experiences. These memories, can be recalled at short notice and used to make predictions about future events, which can then be used to guide behaviour. The development of similar, synthetic memory systems will endow robots with the ability to store their previous experience and use this to make inferences about current and future events. These inferences can then be used by the robot to autonomously adapt its behaviours to deal with different situations and environments. This will form the basis for a more complete social interaction.

Human memory performs complex processes, encoding, retrieving and storing information from multiple sensory streams, for example an object can be remembered from its properties, which could include appearance, smell, taste and texture. The formation of memories therefore requires a fusion of multiple sensory inputs. Thus, the design and development of robust synthetic memory systems should replicate multi-sensory fusion of information from all the available sensory modalities on the robot platform, for example odometry, touch, vision and sound.

As a first step the multiple sensory inputs of the Pioneer robot will be used to generate memories during exploration of a local environment. The robot will then navigate around the same space using both the built-in mapping and synthetic memory systems. The combination of these systems is expected to improve the navigational accuracy of the robot, especially in dynamically changing and novel locations, as the current sensory information can be used to recall memories and increase location inference. A further development will utilise the synthetic memory system, with visual and audio input and output, to recognise and interact with humans during exploration and navigation. This will provide a foundation for the development of human-robot interaction based on bio-inspired memory storage and recall.

3 Preliminary experiment

In the remaining sections we present a preliminary investigation into the ability of the robot to identify which floor of the building it is on[3] , using information

[3] Our experiments are based around the lift in the Pam Liversidge Building at The University of Sheffield. Floors are alphabetically labelled, from the ground up, and include floor 'C+' due to a neighbouring mezzanine level

readily available to it. The data comes from three sources: an on-board microphone, which can detect announcements made within the lift; visual inspection of the floor signage outside the lift; and map data previously collected from each floor of the building. These three sources are available to the robot at different times: whilst travelling within the lift; once the lift doors have opened, or from directly outside the lift; or along the corridor away from the lift.

We analyse the ability of the robot to identify which floor it is on using each source alone, then combine the measures using a Bayesian filter and assess whether these available measures are sufficient for floor identification.

3.1 Vision

There is an information panel outside the lift on each floor of the Pam Liversidge Building, shown to the left of Fig. 1. This panel indicates which floor the viewer is currently on, and can be used by the robot to determine its location in order to load the correct floor map. Due to the distance of the panel from the lift, and the relatively low camera resolution, it is not possible for the robot to read the text via an Optical Character Recognition (OCR) technique whilst still within the lift. Instead, the distance between two blue bars, indicating the building and current floor, is correlated with each floor.

Fig. 1 illustrates the process of floor detection based on an image of the information panel. Firstly, the image is thresholded in the Hue-Saturation-Value (HSV) colourspace to isolate the blue areas as a binary image. This thresholded image is then passed to a blob-analysis routine which calculates the area and centroid position of the two largest blobs. The distance between the blobs is then calculated as

$$d_{scaled} = \frac{d_{centroid}}{\sqrt{A_{largest}}} \tag{1}$$

where $d_{centroid}$ is the Euclidean distance between the centroids and $A_{largest}$ is the area of the largest blob. This scaling ensures the algorithm is insensitive to the size of the information panel in a given image.

Finally, the scaled distance is compared with reference values for each floor to produce a Probability Mass Function (PMF)

$$P_i = \frac{1}{\rho}|d_i - d_{scaled}|^{-1} \tag{2}$$

where d_i is the reference distance of the ith floors information board and ρ is a normalisation term given by

$$\rho = \sum_{i=1}^{N} |d_i - d_{scaled}|^{-1} \tag{3}$$

where N is the number of floors.

Fig. 2 illustrates the PMFs for the information panel on each floor. It can be seen that floors B-E are clearly distinguishable, however floors A and F are

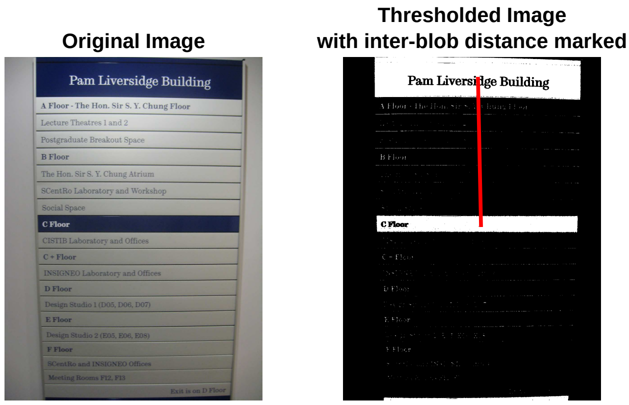

Fig. 1. Detection of the C floor information panel

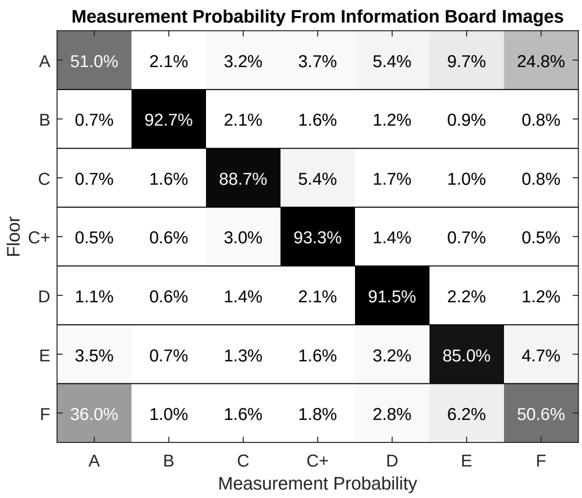

Fig. 2. PMFs for each floor information panel

easily confused. This confusion is explained by Fig. 3, where it is clear that the blue bar highlighting floor A is not distinct from that indicating the building, therefore the second blue blob used by the analysis is that at the very bottom of the panel. This produces a d_{scaled} value similar to that for floor F, leading to this confusion.

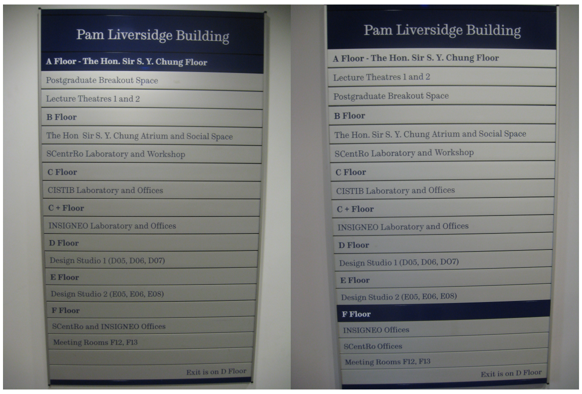

Fig. 3. Information panels for floors A and F

3.2 ASR of floor announcements

We collected audio in the lift using a laptop and a low-budget USB far-field microphone, the Andrea External USB Soundcard/SuperBeam Microphone Bundle[4], to simulate the data that might be collected by the Pioneer in real operations. In total we collected 9 minutes and 35 seconds of audio over three sessions. During each session we travelled up and down all the floors to hear the range of announcements that were given by the automated lift speaker. The data was manually segmented and transcribed to separate each lift utterance from the rest of the background sound and other noises.

The data was analysed using our speech recognition system. We used a cross-validation approach to evaluate the output of the recogniser, using each combination of two sessions as training data and the third as the test. The recognition system was built using the Kaldi toolkit [11], using the SGMM decoding approach [12]. The acoustic models were trained on the WSJ British English spoken corpus. We used the speaker adaptation (SAT) scripts to adapt the models to the acoustic conditions of the lift. The pronunciation dictionary was designed to fit the phrases uttered by the lift, and the language model was created as a constrained grammar to only allow the phrases that are uttered by the lift.

In total there were 20 direction and 20 floor announcements. For the direction announcements 17 were identified correctly, giving an accuracy of 85%. For the floor announcements 11 were identified correctly, an accuracy of 55%. For each announcement we obtained the n-best paths output from the recogniser. We looked at the acoustic model scores output for each path to estimate the confidence distribution for the floors or directions, and converted these into PMFs. The PMFs for one set of announcements are shown in Fig. 4. In many cases the values are very close, due to the level of background noise; this indicates the high level of potential confusion in the floor identification.

[4] `http://www.andreaelectronics.com/`

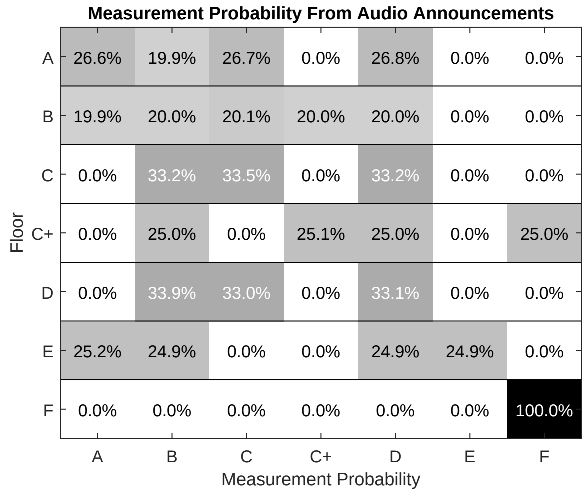

Fig. 4. PMFs for each floor announcement

3.3 Mapping

Before the experiment, we used the onboard laser scanner and inbuilt software (MobileEyes, Mapper 3, and ARNL Server) to map the 7 floors of the building. The floorplan of the area directly outside the lift is identical on every floor, and so it is necessary for the robot to leave this area before it can detect differences in layout afforded by the various floors.

For our experiment, the robot was driven to a point just outside the lift lobby on each floor, where it was able to detect the layout of the adjacent corridor. A measure of confidence as to which floor the robot was on was then generated using MobileEyes by comparing the generated laser point cloud with the pre-recorded maps of each floor. This data was converted into PMFs in a similar way to section 3.1, and the results are given in Fig. 5. Floors C, C+, D, and E are similar in their layout, and this is reflected in the data.

3.4 Combined floor estimation

The previous sections have detailed processes for obtaining PMFs from various sensors, indicating the confidence of the robot being on a particular floor. These measurements are combined using a Bayesian filter to produce a final estimate of the robots position. To improve this estimate it is assumed that the robot knows the floor it is starting from precisely and is able to detect whether the direction of the lift from the announcements with 85% confidence (i.e. there is a 15% chance the robot will think it is going down when it is in fact going up).

Fig. 6 illustrates the execution of the Bayesian filter based on the data presented in the previous sections. It can be seen that the prior distribution indicates the robot is 100% confident in its location (D Floor). After the lift begins

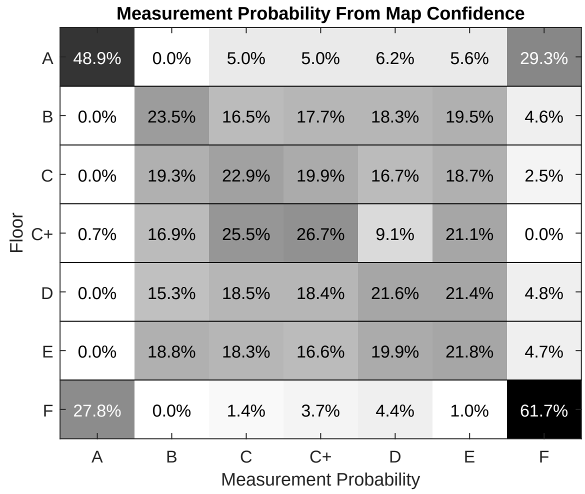

Fig. 5. PMFs for each floor map

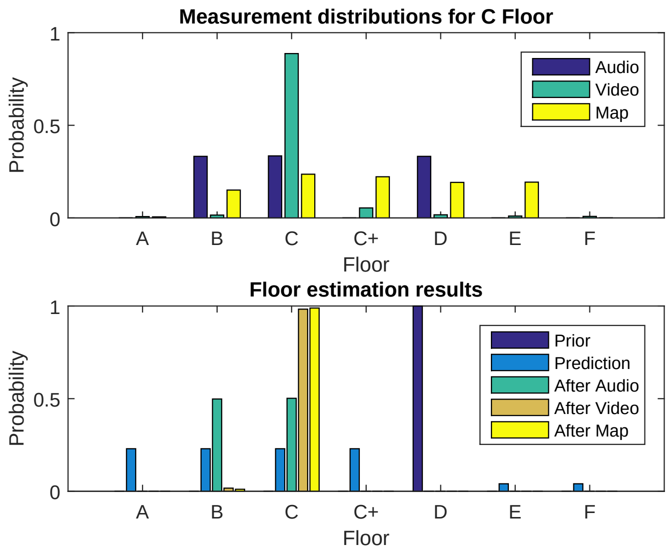

Fig. 6. Prediction of the current floor by the robot. Initially located on D floor, travelling to C floor

to move, the robots prediction of which floor it will leave on is only based on the direction, leading to an indistinct distribution.

Once the lift stops and announces the floor, the robot is unable to precisely distinguish the announcement between floors B, C and D. Combining this measurement distribution with the prediction, however, discounts D floor as the robot knows this is where it started.

After leaving the lift and capturing an image of the information board, a more distinct measurement is obtained and combined with the estimation. Finally, the robot begins to drive around the floor and assesses the validity of each floor map, increasing its confidence further. Comparison of the individual measurements with the final estimated distribution shows that the incorporation of the additional measures has increased the robots confidence of being on Floor C from 89% (based on the video measurement) to over 98%.

Floor	A	B	C	C+	D	E	F
Best Measurement	51.1%	92.7%	88.7%	93.3%	91.5%	85.0%	100%
Combined Estimate	99.7%	97.2%	98.3%	99.4%	98.3%	99.9%	100%

Table 1. Comparison of the best measurement confidence for each floor with the estimate achieved by combining measurements

Table 1 illustrates how the floor estimation confidence compares with the best measurement for each floor. With the exception of F Floor, which is detected perfectly by the audio measurement, all floors show a significant confidence improvement by aggregating multiple measures.

4 Conclusion

In this paper we have introduced our interdisciplinary approach to the development of a guide robot, and highlighted the dependencies between the various areas of research. We have also described an initial investigation into navigation between floors in a building, and our approach to floor identification based on the fusion of readily-available indicators.

Our results show a range of success in using audio, visual, and laser data to correctly identify the floor on which the robot is situated. In almost every case, we have also shown how the Bayesian filter applied to all measures improves this estimation.

In this work, measurements were taken in near ideal conditions, without other building users interfering with data collection. This is not to be expected in standard operation, and we are now extending the experiment described here to investigate a broad range of realistic operating conditions.

References

1. Aitken, J.M., McAree, O., Boorman, L., Cameron, D., Chua, A., Collins, E., Fernando, S., Law, J., Martinez-Hernandez, U.: Safety and verification for a mobile guide robot. In: Proceedings of Towards Autonomous Robotic Systems (TAROS) (submitted) (2015)

2. Anderson, T., Borah, J., Leger, A., McMillan, G., Rood, G.: The Technology of Speech-Based Control. Alternative Control Technologies: Human Factors Issues (1998)
3. Cameron, D., Collins, E.C., Chua, A., Fernando, S., McAree, O., Martinez-Hernandez, U., Aitken, J.M., Boorman, L., Law, J.: Help! I can't reach the buttons: Facilitating helping behaviors towards robots. In: Biomimetic and Biohybrid Systems. p. In Press. Springer (2015)
4. Clark, R.D., Word, L.E.: Why don't bystanders help? because of ambiguity? Journal of Personality and Social Psychology 24(3), 392 (1972)
5. Dixon, C., Webster, M., Saunders, J., Fisher, M., Dautenhahn, K.: the fridge door is opentemporal verification of a robotic assistants behaviours. In: Mistry, M., Leonardis, A., Witkowski, M., Melhuish, C. (eds.) Advances in Autonomous Robotics Systems, Lecture Notes in Computer Science, vol. 8717, pp. 97–108. Springer International Publishing (2014)
6. Hancock, P.A., Billings, D.R., Schaefer, K.E., Chen, J.Y., De Visser, E.J., Parasuraman, R.: A meta-analysis of factors affecting trust in human-robot interaction. Human Factors: The Journal of the Human Factors and Ergonomics Society 53(5), 517–527 (2011)
7. Kelly, T., Weaver, R.: The goal structuring notation–a safety argument notation. Proceedings of the dependable systems and networks 2004 workshop on assurance cases (2004)
8. Kulyukin, V.A.: On natural language dialogue with assistive robots. In: Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction. pp. 164–171. HRI '06, ACM, New York, NY, USA (2006)
9. McAllister, D.J.: Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations. Academy of management journal 38(1), 24–59 (1995)
10. MoD Interim Defence Standard: Standard 00-56 issue 4. Safety Management Requirements for Defence Systems (2007)
11. Povey, D., Burget, L., Agarwal, M., Akyazi, P., Feng, K., Ghoshal, A., Glembek, O., Goel, N.K., Karafiát, M., Rastrow, A., et al.: Subspace gaussian mixture models for speech recognition. In: Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on. pp. 4330–4333. IEEE (2010)
12. Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., Vesely, K.: The kaldi speech recognition toolkit. In: IEEE 2011 Workshop on Automatic Speech Recognition and Understanding. IEEE Signal Processing Society (Dec 2011)
13. Rosenthal, S., Biswas, J., Veloso, M.: An effective personal mobile robot agent through symbiotic human-robot interaction. In: International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010). vol. 1, pp. 915–922 (May 2010)